

Comparison of Methods for Evaluation of Medical Terminological Systems

D.G.T. Arts, MSc^{1,2}, R. Cornet, MSc¹, E de Jonge², PhD, N.F. de Keizer, PhD¹

¹ Dept. of Medical Informatics, Academic Medical Center, Amsterdam, The Netherlands

² Dept. of Intensive Care, Academic Medical Center, Amsterdam, The Netherlands

ABSTRACT

The importance of terminological systems (TS) for the medical domain is widely recognized. The usability of such a system depends primarily on its content. We have designed four methods to evaluate the content of TS and applied them in a case study.

INTRODUCTION

Several developments in health care have led to an increase in the need for accurate, detailed and structured registration of medical data. To support standardized and structured registration of patient information many terminological systems (TS) have been and are still being developed. A TS relates concepts and terms belonging to a specific domain¹. The aspects of TS that can be evaluated and the methods that can be applied are various. This study provides a comparison of four methods for evaluation of the content of TS by applying them in a case study on a TS for the domain of intensive care, DICE (Diagnoses for Intensive Care Evaluation).

METHODS

As in most recent publications examining TS we will focus on the evaluation of the content of TS. 'Content' here refers to the concepts that are included in the TS, their terms and (hierarchical and non-hierarchical) relations. We have designed four methods for the evaluation of the domain completeness and the correctness of the content of TS and applied them in a case study on DICE.

Method A: Field testing

The TS is used in real practice by the intended users of the system, e.g. physicians, to register patient information. The user assigns a 'match score' to each coded concept, indicating the extent to which it matches the concept he has in mind.

Method B: Research queries

A set of concepts, e.g. clustering criteria, is obtained from studies that are related to the domain of the TS. The concepts are given a 'match score' indicating the extent to which they are represented in the content and structure of the TS.

Method C: Description logics

The semantics of the TS are represented using a description logic (DL) formalism. Reasoning with the knowledge in DL reveals inconsistencies that lead to the uncovering of missing or incorrect relations.

Method D: Expert knowledge

Experts in the domain of the TS manually review the concepts and their relations and terms.

EXPERIENCES

Table 1 displays the extent to which each method uncovered incompleteness or incorrectness in the content of the TS. Methods A and B differ only in the way of retrieving the concepts that are being matched to the TS. Method B appeared to be more suitable to retrieve rare concepts. Disadvantage of method B was that clustering criteria are often vaguely described. It appeared to be a time consuming effort to uncover the causes of the inconsistencies that were identified by method C. However this method has the potential to support maintenance and evaluation of TS through an automated approach. Method D, although time consuming, appeared to be very valuable, since concepts that were included in this TS had not yet been evaluated by an editorial board.

Table 1 – The extent to which each method identified the types of incompleteness and incorrectness.

		A	B	C	D
Incomplete	Concepts	+	+	-	+/-
	Terms	+/-	+/-	-	+/-
	Non-hierarchical rel.	+/-	+/-	+	+
	Hierarchical rel.	-	-	+	+/-
Incorrect	Terms	+/-	+/-	-	+
	Non-hierarchical rel.	-	-	+	+
	Hierarchical rel.	-	-	+	+

CONCLUSION

All four methods appear to be valuable and feasible for evaluating the content of TS. Which method to apply depends on the focus of the evaluation. Table 1 can be a useful reference. Research on other issues, such as user interface and maintenance is ongoing.

REFERENCES

1. de Keizer NF, Abu-Hanna A, Zwetsloot-Schonk JH. Understanding terminological systems I: Terminology and typology. *Methods Inf Med.* 2000 Mar;39(1):16-21.

This study was supported by the Dutch Intensive Care Evaluation (NICE) foundation. (www.stichting-nice.nl)